DOCUMENT RESUME

ED 135 849 TM 006 079

TITLE Evaluation and Analysis to Support Decisionmaking.
INSTITUTION Comptroller General of the U.S., Washington, D.C.
REPORT NO PAD-76-9
PUB DATE 1 Sep 76
NOTE 55p.

EDRS PRICE MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS *Decision Making; *Evaluation; *Evaluation Methods; Federal Government; *Federal Programs; Guides; *Policy; *Program Evaluation; Resource Allocations

ABSTRACT

Legislation requires that the Comptroller General of the U.S. develop and recommend to the Congress methods for review and evaluation of government programs carried on under existing law. This document is a first step in collecting and disseminating general concepts on these activities and how they are related to other activities in the continuum of decisionmaking about government programs. The document is offered for the use of anyone who is evaluating programs and analyzing policy alternatives in the sense of engaging in a careful appraisal of what happened, why it happened, what choices are available for future actions and what the implications are of those choices. It is intended for use by all persons who are concerned with this process whatever the academic discipline or professional background from which they approach the problem. While addressed primarily to practitioners, it is hoped that this publication will also be a useful reference to those who, as legislators or managers, for example, are interested in the products of evaluation and analysis. The chapters discuss the framework within which these activities are performed and provide both conceptual and practical guidance. The emphasis throughout is not on advanced quantitative techniques, but on essential concepts and basic approaches. (EC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

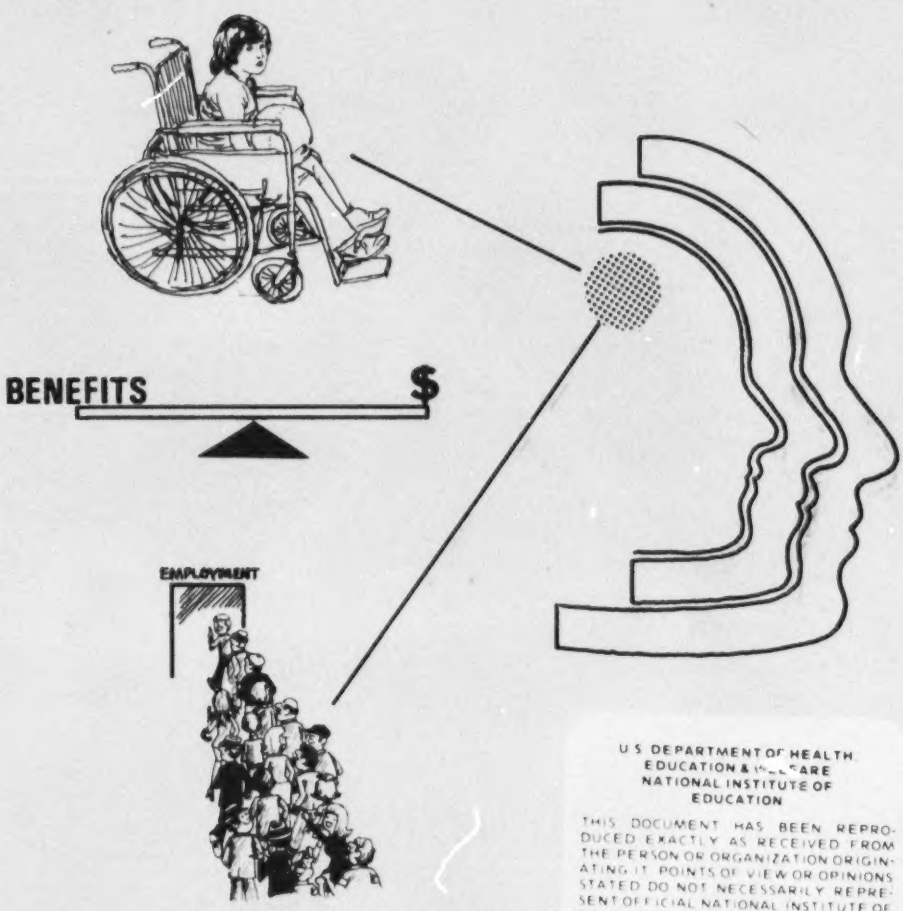
ED135849

Requested

EVALUATION AND ANALYSIS TO SUPPORT DECISIONMAKING

unsolicited

TM



U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

UNITED STATES GENERAL ACCOUNTING OFFICE

TM006 079

PAD-76-9
REVISION OF OPA-76-9 DECEMBER 1975

SEPT. 1,1976

B-177913

For a number of years, we have been expanding the scope of our work and the capability of our staff to review the results and effectiveness of Government programs.

The Comptroller General is given very broad responsibility and authority under the Budget and Accounting Act, 1921, and the Accounting and Auditing Act of 1950 for the evaluation and analysis of Federal programs and activities. Section 204 of the Legislative Reorganization Act of 1970 supplemented our authority and indicated current congressional interest in analyses of programs. The Congressional Budget Act of 1974, further strengthened the congressional emphasis on the evaluation of programs, the statement of legislative objectives and goals, and improved methods of evaluation. Such work is carried out by all parts of GAO in connection with our general review, evaluation, analysis, and audit functions.

The 1974 act requires, among other things, that "the Comptroller General shall develop and recommend to the Congress methods for review and evaluation of government programs carried on under existing law." This document is a first step in collecting and disseminating general concepts on these activities and how they are related to other activities in the continuum of decisionmaking about Government programs. The document adds to guidance contained in Standards for Audit of Governmental Organizations, Programs, Activities and Functions, issued in 1972, which includes in the full scope of such audits a review to determine whether desired results are effectively achieved.

The review and evaluation of programs and the analysis of alternatives are central to our mission. These activities are also critically important to effective management in executive agencies. An agency with the capacity to assess the impact of its programs and to examine alternative courses of action is much more likely to pursue its program goals effectively. One objective of this document, therefore, is to encourage agencies to develop--and use effectively--this capacity for critical self-assessment.

This document is intended to be of value to the novice and the experienced practitioner whether engaged in financial audits, program review, or in program evaluation or analysis.

It is not a substitute for the many good textbooks available on particular analytical methods. Rather, it establishes a conceptual framework within which analysts from varying backgrounds can work and communicate effectively.

We received numerous helpful comments and suggestions on an earlier draft of this document. We have taken those comments into consideration in preparing this final version.

We trust that this document will serve its intended purpose, to encourage more effective use of evaluation and anaylsis in decisionmaking.

Elmer B. Staats

Comptroller General
of the United States

Contents

CHAPTER		Page
1	INTRODUCTION	1
2	THE PUBLIC DECISIONMAKING PROCESS	3
	The resource allocation problem	4
	Issues in resource allocation	4
3	THE EVALUATION AND ANALYSIS CONTINUUM	6
	Concurrent processes in the continuum	6
	Roles people play	9
4	APPRAISING THE RESULTS OF POLICIES AND PROGRAMS AND ASSESSING ALTERNATIVE APPROACHES	11
	Understanding fundamentals	11
	Ascertaining decisionmakers' needs	11
	Defining the nature and scope of the problem	13
	Determining valid objectives	13
	Specifying comprehensive measures	15
	Appraising policy and program results	16
	Making valid comparisons	17
	Developing needed information	20
	Interpreting program results	21
	Checking completeness of the appraisal	22
	Assessing policy and program alternatives	23
	Developing a range of alternatives	23
	Screening the preliminary alternatives	24
	Estimating measurable consequences	25
	Assessing provisional orderings	26
	Determining the impact of constraints	27
	Reassessing the orderings of the alternatives	28
	Checking completeness of the assessment	29

CHAPTER		Page
5	PRACTICAL ASPECTS OF MANAGING AND PERFORMING STUDIES	31
	Formulating an agenda of studies	31
	Identifying emerging problems	31
	Deciding which problems to study	32
	Beginning a study	33
	Preparing a detailed study plan	33
	Selecting the study team	34
	Establishing lines of communication	35
	Selecting appropriate methods	35
	Conducting a study	37
	Collecting relevant data	37
	Testing the reliability of data	37
	Protecting the confidentiality of information about individuals	38
	Documenting and referencing	39
	Adhering to time schedules	40
	Leading and coordinating the study team	40
	Using computer-based models	40
	Communicating study results	41
	Specifying the nature of reports	41
	Communicating with clarity and conciseness	42
	Following up	43
APPENDIX		
I	An annotated bibliography	44

CHAPTER 1

INTRODUCTION

In recent years, a multitude of new programs have been launched to deal with the problems of our society, and Government expenditures have increased at a very rapid rate. Are the programs working? Are the funds being spent on these programs producing the results desired? Are there better ways to attack and solve society's problems? And, are old programs achieving their objectives?

Governments and agencies entrusted with public resources and the authority for applying them have a responsibility to answer these questions--to render a full accounting of their activities. Government managers have a responsibility to show not only the purposes for which public resources were used, but also to demonstrate the effects of their use.

This responsibility rests first on the executive agency administering the program. Program managers need to know what the programs entrusted to them are accomplishing and whether results might be improved. Central management agencies--the Office of Management and Budget or the White House, for example, need to know if programs are working--either through their own efforts or through the review of the appraisals of managing agencies. Finally, congressional responsibility for legislation, appropriations, budgeting and priorities, and oversight and investigations indicates a need for the legislative branch to make its own appraisals of programs and to make use of appraisals made by the executive branch.

The ultimate choices about programs--decisions about whether to do or not to do something--will be policy choices. However, political leaders, public administrators, and the public need as much information as possible on the choices that must be made. This need has stimulated the development of various analytical techniques which have been grouped under labels such as program evaluation and policy analysis. The art of evaluation and analysis is not yet sufficiently developed to permit preparation of a manual covering "how to do it" in every situation. It is only evolving. While recognizing this, decisions must and will be made by legislators and executives faced with the task of formulating and reformulating programs to deal with the problems of our society.

Thus, we offer this document as a first step in collecting

and disseminating lessons learned in GAO and elsewhere about analysis and evaluation. Generally speaking, we offer this guidance for the use of anyone who is "evaluating" programs and "analyzing" policy choices in the sense of engaging in a careful appraisal of what happened, why it happened, what choices are available for future actions, and what the implications are of those choices. We intend it for use by all persons who are concerned with this process whatever the academic discipline or professional background from which they approach the problem. While this statement is addressed primarily to practitioners, we hope it will also be a useful reference document to those who, as legislators or managers, for example, are interested in the products of evaluation and analysis.

The concepts and guidance which we offer must be adapted to specific program situations. Program objectives are seldom as clearly stated or agreed upon as would be desirable for evaluative purposes; no program operates in isolation from other social or economic events; and data and measurement techniques are almost always less adequate than desired. It is in the adaptation of the ideal and the theory to the specific situation that the persons doing the work show their worth. The judgments involved in the identification of objectives, the selection of data and measurement techniques, and the evaluation of external factors transcend in importance any "rules" which can be prescribed in a document of this sort.

The chapters which follow discuss the framework within which these activities are performed and provide both conceptual and practical guidance. The emphasis throughout is not on advanced quantitative techniques, but on essential concepts and basic approaches.

CHAPTER 2

THE PUBLIC DECISIONMAKING PROCESS

There are many reasons for governmental activities. For example, national defense is provided because there is no suitable private alternative. The provision of free public education reflects society's preferences and its notions about its own long-term welfare. Health and safety considerations lead to the regulation of private activities to avoid adversely affecting the public's well-being. Tax incentives such as the investment tax credit are intended to affect the performance and growth of the economy. The relevance and effectiveness of any governmental activity may be questioned. Analysis, evaluation, and related activities are tools to help decisionmakers in resolving these questions.

The private marketplace has limitations, particularly because it does not produce certain goods with high social value. It does not always provide adequate information, sufficient competition, efficient designs or qualities of certain goods, desirable distribution of income and wealth, or desirable modifications of consumption patterns. But, resorting to governmental action because of these failings does not automatically insure that the same or other failings will not occur. The effectiveness and efficiency with which the Government performs its functions must also be weighed in deciding whether governmental action should be enhanced, changed, or in some cases, is warranted at all.

Each year as legislative, budgetary, and appropriations decisions are being considered, the practical issue remains: What does the public need and how should priorities be established? In a democracy, the political process is relied upon to examine and determine public need and to set priorities as to how such needs are to be met from public funds. Elected officials are responsible for learning and reflecting their constituents' needs and proposing programs or program changes with requisite funding levels to assist in determining priorities for action.

But for the work of elected officials to have meaning, accurate and relevant information must be available and useful debate must take place. Analysis and evaluation help provide the needed information and a basis for judgment of those persons and groups involved in public decisionmaking.

THE RESOURCE ALLOCATION PROBLEM

Government actions generate benefits and incur costs. These benefits and costs should be broadly defined to include their social and private aspects. The key elements of the problem of choice are:

--Government objectives are achieved by developing, adopting, and implementing policies and by creating and operating programs, all of which consume or transfer resources--tangible and intangible.

--There are many public needs. These needs are large and constantly changing. Demands for resources are much greater than the resources available.

--Decisionmakers must choose among competing objectives and among the alternative programs and policies capable of meeting the chosen objectives at desired and affordable levels of achievement.

Thus, decisionmakers are involved in the process of allocating available resources among competing demands so as to achieve the greatest overall level of net benefits possible. At the same time, full consideration is given to justice, equity, and political reality.

ISSUES IN RESOURCE ALLOCATION

Most programs are interdependent and affect more than one of society's goals. This leads to a need for use of evaluation and analysis in two resource allocation issues: (1) choices within a major program area and (2) choices among major program areas.

For choices within a program area, the following questions are often posed:

--What is the appropriate level of attainment for a given objective?

--Are there preferred alternatives for reaching that level?

--What are the resource requirements for doing so?

--Are there obstacles to acceptance and implementation of an otherwise preferred alternative and what would be the costs of overcoming the obstacles?

--Are there equity considerations connected with the leading alternatives?

For choices among major program areas, similar questions are relevant. Judgments that are made by decisionmakers concerning relative importance of the various objectives will affect the assignment of resources among those objectives.

Resolution of resource allocation issues may result in any of a number of actions: (1)(a) continue, modify, or abandon existing policies and (b) adopt new polices or (2)(a) continue, modify, expand, reduce, or phase out current programs and (b) create new programs.

The decisionmaking process and the relationships among its component functions do not necessarily follow a predetermined sequence. The decision process and the supporting functions of evaluation and analysis are concurrent, continuous processes, with continuing interaction among the various parts.

The role of evaluation and analysis in the allocation of resources is to provide better information about the implications of the choices available to the decisionmaker.

CHAPTER 3

THE EVALUATION AND ANALYSIS CONTINUUM

Evaluation and analysis covers a wide range of activities designed to support the ongoing decisionmaking process. These activities include reviews known as program auditing, budget examination, management analysis, planning, institutional research, program budgeting, systems analysis, engineering economics analysis, program evaluation, policy analysis, cost-benefit analysis, etc.

The specific type of review needed depends, in part, on the particular issues raised or questions asked and the focus of the inquiry on the part of the decisionmaker. If financial information is needed about a Government agency or corporation, an audit of the results of operation and financial condition for a given period may be the appropriate form of review or analysis needed. If there is concern about the management of a program, a management review that is directed more towards the effectiveness of the organization, management, and staffing may be appropriate.

General program reviews that look at program effectiveness and consequences, as well as management effectiveness, are useful to the decisionmaker in determining whether the program is meeting the established objectives and whether there are changes needed to improve the program efficiency and effectiveness. A comprehensive review of an existing program may consider the overall performance of the program, including an evaluation and analysis of performance on any number of criteria. Consideration of alternatives may also be included.

It should be recognized that traditional perceptions of terms such as "evaluation" and "analysis" tend to overlap. In some cases, the two terms are used interchangeably. Other terms, such as "auditing" or "planning," may also be used interchangeably with evaluation and analysis.

CONCURRENT PROCESSES IN THE CONTINUUM

For purposes of this document, drawing sharp distinctions between evaluation and analysis is less useful than focusing on the two basic questions which decisionmakers, and their staffs, face: (1) What actually has happened as a result of past or current policies and programs and what have we learned? and (2) What should be done in the future and

what are our options? Answering these questions can, in turn, be roughly translated into broad classes of activities: appraising the results of policies and programs and assessing alternative policies and programs. These broad activities include the variety of review types listed previously.

An example of the interaction of these concurrent processes in the continuum is shown in the following illustration.

After a problem is identified, the first step should be a systematic attempt to assess what caused the problem and what the alternative ways are to solve it. Alternative approaches, developed as part of the assessment, do not normally contain the level of detail that would be needed for actual implementation. For example, after the decision is made to implement a program, the initiation of the program would require a detailed plan that specified some or all of the following: processes and technologies to be employed, investment and operating funds needed, and capital facilities needed together with the acquisition schedule that will provide for the scale of operations desired within the appropriate timeframe.

It is important to build into the implementation plan specific provisions for gathering information necessary for a comprehensive and valid appraisal of results. During the implementation phase, such information should include, where possible, the effects of any changes in the implementation plan.

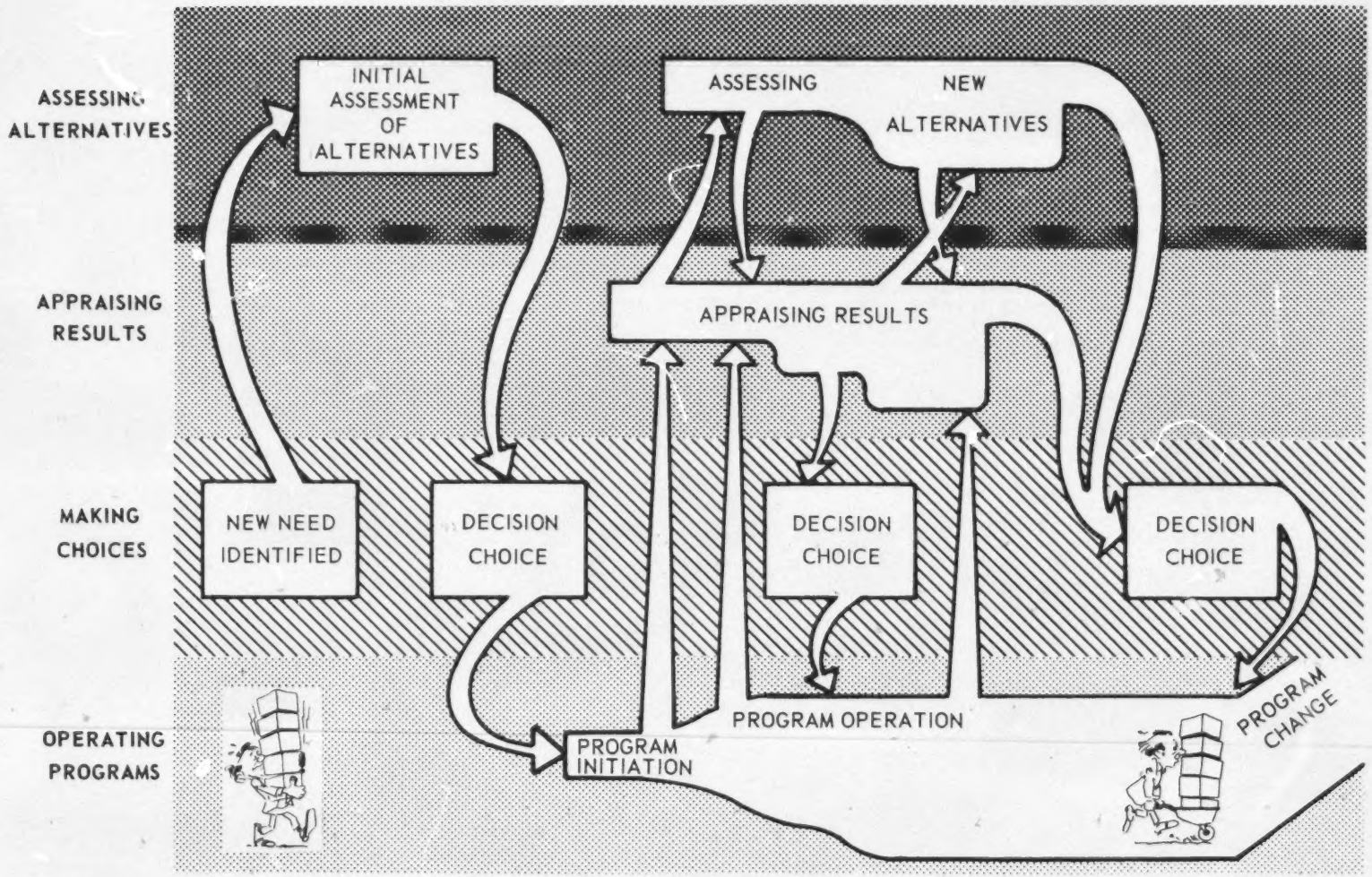
As required to support decisionmaking, further assessment of alternatives should take place concurrently with appraisal of results during the operation of the program. Appraisal of results without such assessment provides only limited insights since it does not address the desirability of new alternatives. Conversely, assessment of alternatives without sound appraisal of results lacks the insights which would come with a full understanding of the existing program. A high degree of interaction exists not only within the various phases of an assessment or appraisal, but between the two processes. If these interactions and feedbacks are ignored, the quality of both efforts suffers.

The combination of appraising results and assessing alternatives should answer the following questions:

--What was the impact of the existing program?

--Could a comparable level of results, effectiveness, or benefits be achieved at lesser cost;

EXAMPLE OF THE CONCURRENT PROCESSES IN THE CONTINUUM (FOR A NEW NEED AND A NEW PROGRAM)



or for the same costs could greater effectiveness, or benefits, be realized?

--Could these more desirable outcomes be achieved by reshaping or redirecting the current program or policy, adopting some previously suggested, but rejected, alternative; substituting some wholly new alternative; or creating some combination of the foregoing possibilities?

Real world decisionmaking and many of the activities which support it are complex, uncertain, hurried, and subject to all kinds of constraints: some understandable and some seemingly arbitrary. These difficulties are compounded by the existence of competing or complimentary objectives (or their related programs) which may also have to be considered. Evaluators, analysts, and other reviewers can do little about these difficulties; but, if they understand the complexity of the situation, they can perform in a way which is of maximum usefulness under the circumstances.

ROLES PEOPLE PLAY

The distinction between the concurrent activities on the one hand and the people or organizations that perform them on the other hand is an additional source of confusion. As noted earlier, appraising policy and program results and assessing alternatives are related and mutually reinforcing processes. These processes can be, and frequently are, performed within a single organization and often by the same person. This has the significant advantages of efficiency and of keeping the practitioners of the various skills aware of useful interactions.

In the real world situation a variety of people are involved who have different backgrounds and call themselves by various titles. Many of these people move between appraising the results of policies and programs and assessing the alternatives for improved choices in the future. At times, these people may be emphasizing the assessment of alternatives, at other times the appraisal of results, and sometimes both together. People who gain their first experience in one activity move throughout the continuum and interact with other people having other experience when working on a particular study.

No group or profession has a monopoly on the talents required of a good evaluator or analyst. The basic prerequisites are (1) an inquiring, skeptical, challenging mind, (2) the ability to think systematically and rigorously, and (3) an openness to new ideas. This "mind set" obviously needs to be coupled with an appreciation of the uses, powers, and limitations of such fields as economics, statistics, accounting, operations research, etc. When high levels of skill in these and other areas are required, the practicing evaluator, analyst, or other reviewer should recognize this fact and call on the needed expert.

The following chapters focus on those ideas, concepts, and approaches which are basic to appraisals of policy and program results and to more insightful assessments of alternatives for improvements.

CHAPTER 4

APPRAISING THE RESULTS OF POLICIES AND PROGRAMS AND ASSESSING ALTERNATIVE APPROACHES

The previous chapters discussed problems and issues in public decisionmaking with particular emphasis on resource allocation and on the continuum of evaluation, analysis, and other review functions which support decisionmaking. This chapter discusses the methods and concepts associated with evaluation and analysis.

The discussion that follows focuses on

--understanding fundamentals in appraising results and assessing alternatives of programs and policies,

--appraising policy and program results, and

--assessing policy and program alternatives.

Obviously, the degree to which the methods can be applied in a particular case depends on the specific problems to be considered. It is not always necessary to use each of these methods in the order presented here.

UNDERSTANDING FUNDAMENTALS

The activities of appraising results and assessing alternatives of programs and policies share certain fundamental concepts in which the mode of inquiry is essentially the same. These fundamental concepts include

--ascertaining decisionmakers' needs,

--defining the nature and scope of the problem,

--determining valid objectives, and

--specifying comprehensive measures.

Ascertaining decisionmakers' needs

An initial task in either appraising results or assessing alternatives is to develop a clear understanding of the decisionmaking needs. These needs can generally be summarized by answering a series of questions.

--What is the decisionmaker's perception of the problem?

--Is there dissatisfaction with effectiveness or consequences of the policy or program? With the lack of a policy or program?

--What uses are to be made of the information to be collected?

--When is the final report needed? Will interim reports be needed?

It is helpful to make some distinctions among the various participants in the decisionmaking process. Some may already have an understanding of the nature of the problem they face, what they want to know, and why. Others may only have a general perception of the problem and what needs to be done about it. In the latter case, a more extensive discussion of these fundamentals may be needed to develop the basis for a study that will be useful to these decisionmakers. The nature of the issue will have a major impact on the sort of information that must be collected and the sort of analysis that must be done. Some environmental questions, for example, are not as susceptible to precise answers as would be desired by the decisionmaker.

It is important to recognize different (and possibly conflicting) viewpoints and interests among participants in the decisionmaking process. The official sponsor may be a congressional committee, whereas the real user of the study may be one member of the committee or the committee staff. Other participants in the process include the managers of the program being evaluated or analyzed, as well as managers of related programs. Another congressional committee, the Office of Management and Budget, or a private organization, such as an association or Government contractor, may become increasingly interested as the study progresses.

In developing a clearer idea of any of these participants' needs, attempts should be made to elicit and clarify information on the nature of the problem or issue as it is currently understood, the general context of the problem, and parts that appear to require special emphasis. Specific attention should be given to the order of priority in approaching the various parts of the problem and to particular points of information or insight essential to making the decision or meeting the decisionmaker's needs. The bureaucratic or political context in which the decision will be made needs to be

understood. The time available for the study and the critical points at which specific items of information are needed should be ascertained.

Defining the nature and scope of the problem

It is essential that the intended users of the study (the participants in the decisionmaking process), together with those persons responsible for the study, share a common understanding of the nature and scope of the issues at stake.

A full and correct understanding of the nature of the problem will be aided by (1) considering its origin, if known, (2) reviewing legislative hearings, reports, and acts associated with it, (3) inquiring into the history of programs designed to deal with the problem, and (4) examining past analyses, evaluations, audits, and budget examinations of the same or related issues. Outlines and checklists can be helpful in this review.

The scope of a study depends both on the questions it would be desirable to answer and the availability of methods and data which will provide those answers. There is often a tradeoff between the breadth of a study and precision in the results. The planned scope should also consider the nature of the decisions which the study may affect. For example, a broad consensus that a program is not working may warrant paying more attention to basic alternatives and less to minor improvements in the existing program. A shared understanding of the scope of the study and the objectives and measures of the policy or program are the foundation for defining the initial direction of the study effort.

Similarly, understanding is needed of the coverage required in geographical terms (regional, State, local); on areas, populations, individuals, or units to be included; and on the scope (how many individuals, approximately how much information from each, etc.). The scope of coverage together with the timing provides for the logistics of the work. For example, indepth work may be undertaken at a small number of locations, less detailed work at a larger number of locations, or some combination of these.

Determining valid objectives

The objectives of the policy or program--the benefits desired to be achieved--frequently are not stated clearly and precisely. The original sponsors of the policy or program

may not have had a precise idea of the end results desired. Formal statements of objectives may be intentionally ambiguous if it is easier to obtain a consensus on action. Value judgments underlying the objectives may not be shared by important groups. Consequently, the end results intended may be perceived by some as implying ill effects for them. Furthermore, explicit statements of objectives tend to imply a specific assignment of priorities and commitment of resources.

To the extent feasible, statements of objectives should:

1. Capture a complete understanding of the intended benefits, including the expected level of attainment.

2. Identify recipients of unavoidable adverse consequences or unintended benefits.

3. Include important qualitative aspects, even though measuring degrees of attainment may be exceedingly difficult.

4. Take account of multiple objectives which may be complementary or conflicting.

The importance of taking such a comprehensive view of objectives cannot be overstated. Oversimplified statements (1) will not capture all essential aspects of the effects intended and (2) may contain implied conflicting consequences for groups other than the intended beneficiaries (e.g., to eliminate hunger or to achieve energy sufficiency). Implied objectives may represent desirable end results. For example, a summer employment program aimed primarily at increasing earnings of young people may be viewed as reducing civil disorders. Moreover, even desired end results may not all be achievable simultaneously and may be interdependent.

Oversimplified statements may result if activity milestones are contained in them (e.g., to increase the number of emergency rooms by 20 percent by 1978). An objective stated in this way may overly constrain an assessment of alternatives, the purpose of which is to determine efficient levels of attainment of an ultimate benefit. On the other hand, it is important that the statement be specific with respect to the nature and direction of change so that progress can be measured. A statement such as "to reduce deaths, additional complications, disability, and suffering of persons with acute injuries by improving emergency care" would satisfy

these criteria. Quantitative goals or targets are also needed, but these must reflect priorities among programs. Accordingly, they can best be set as part of the budget and long range financial planning process and should be reexamined regularly as budget priorities shift.

In appraising results of ongoing programs, if targets or activity milestones have been furnished to managers, the targets or milestones should not be accepted uncritically (e.g., a specified student/teacher ratio). An attempt should be made to find whether deficiencies in attaining the milestones are caused by unrealistic expectations or by the way the program was implemented or operated.

A shift in objectives can occur over time and care must be taken to assure that statements of objectives currently in use are still valid. For example, the objective associated with the national 55 miles per hour speed limit now includes safety, as well as energy conservation.

Determining valid objectives is a complex and frustrating task. A study may have to proceed without fully satisfying these requirements. If this is the case, objectives should be reexamined and clarified as the study progresses.

Specifying comprehensive measures

Valid measures of policy and program consequences are required for both appraising results and assessing alternatives. Objectives and measures of consequences are interdependent. The quality of each depends on the other. Measures should be used which cover all aspects of a given objective. Ideally, measures should

--quantify the extent to which the objective(s) are or would be met--"effectiveness" measures;

--capture qualitative aspects of the consequences--"intangible" measures;

--quantify, to the extent possible, unintended consequences--"side-effect" measures;

--quantify, to the extent possible, the differences of impact on the beneficiaries and the cost bearers--"distribution" measures.

When appraising results, it may be decided for practical

reasons to exclude side-effect and distribution measures. However, they should not be excluded if reasonable effort would produce useful data or if the decisionmaker is especially interested in these measures. For intangible measures, some qualitative indication of relative magnitude should be used (e.g., ratings by clients reflecting their satisfaction with the quality of a service).

Data may not be available on the desired measures or, if available, may be obtainable only at high cost. In these cases, surrogates will have to be used. For example, the scholastic aptitude test is used to measure likely achievement in college. When surrogates are used, their validity should be established.

There is a temptation to define quantifiable measures, especially of effectiveness, too rigidly or narrowly. For example, in evaluating a public employment program, a successful participant might be specified as a person who is employed 1 year after completion of training. If the participant worked 1 day less than a year, would he or she be viewed unsuccessful? Suppose the participant only occasionally held a job, but happened to be working a year after the program. Should this be counted as a success? The range and distribution of outcomes should be considered in this case. For example, data on the percentage of persons holding jobs for various lengths of time after training would provide a more meaningful picture of real outcomes. This sort of distribution is needed to judge levels of attainment or degrees of success or failure.

These four fundamental concepts are closely interrelated. A clear understanding of what is needed for the decisionmaking process, of the nature of the problem, and statements of objectives is necessary in order to assure that a meaningful and feasible set of measures has been specified.

APPRAISING POLICY AND PROGRAM RESULTS

The process of appraising results should begin concurrently with policy or program implementation and continue as needed. Continuous appraisal, through a well structured management information system, should be maintained, but even when it exists there will be a need for special reviews from time to time.

After the fundamental concepts discussed above are understood they must be further developed through application of other more specific concepts and methods including

--making valid comparisons,

--developing needed information,

--interpreting program results, and

--checking the completeness of the appraisal.

Making valid comparisons

Comparison is the essence of appraising program results. For evaluation to be useful, outcomes of the program or policy must be compared with something else in order to reveal the effect of the program. The basis for comparison may be the outcome of an alternative policy, of the same program at an earlier time, or a combination of such comparisons.

The measures that are used in making these comparisons should be derived from a valid statement of objectives. This derivation is more difficult if the objectives are not clearly stated, changes are taking place, or for other reasons the basis of comparison is arbitrary.

The measures used for comparison need to be developed with the decisionmakers who know whether their focus is on resource input (e.g., mixture of paraprofessionals, nurses, and physicians), operational process (e.g., scheduling of surgeries), outcomes (e.g., disability days averted), the operational setting (e.g., interaction with other outpatient and inpatient facilities), or some combination of these.

Other sorts of comparisons may be useful. For example, comparison of planned variations among projects within an existing program may help to identify important characteristics and potential improvements. A comparison of similar programs, if feasible, may provide some of the same information.

Once the nature of comparisons is established, a series of additional questions directly relating to the problem at hand should be raised. Some of these are posed as hypotheses which the appraisal aims to prove or disprove. Such hypotheses or questions must allow an appraisal of whether consequences or effects are attributable to the program or to some other causes. A decision needs to be made whether only descriptive findings will suffice or whether it will be necessary to demonstrate significance of results or differences in effects.

Choice of a comparison approach depends both upon the

questions to be asked and the availability of data. This is not only procedural but involves also questions of access, comparability, restrictions on collection and use of confidential data, etc. These problems may be more severe than many evaluators, auditors, examiners, and others realize. This means that analytical methods should be used which make the most efficient use of each bit of data.

Data problems may make it impossible to use the best theoretical method. Some methods may be impractical if data is too highly aggregated, incomplete or missing, or may require "patch up" efforts after the evaluation is underway.

Some major comparison methods follow.

Experimental methods--attempt to measure the results of the program as though everything else is held constant. This is done by measuring the difference, in terms of the measures of success, between those affected by the program and a control group which is not. This is the preferred method for evaluation of social experiments, but it can also be used for any evaluation when the essential requirements of random assignment and control are feasible. This is the approach that was used in the New Jersey Negative Income Tax experiment. In that experiment, several different amounts of monetary incentives were given to different groups of families in the same situation to see what effect the incentives had on work and spending habits. Responses were compared with the habits of families in the same situations which received no monetary incentive from the experiment during the same time.

The analytical strength of the experimental method makes it a very useful tool. This value must be balanced against other considerations, such as cost and ethical and legal constraints, before this approach is selected.

Experimental designs require that the affected group and the group not affected possess similar characteristics. This is the reason for a strict requirement that the potential participants be randomly assigned so that each one has the same chance of assignment to either group before the program begins. Unless randomization is achieved, there is no assurance that the results are attributable to the program.

For example, unless randomly assigned, participants might enroll because they are more perceptive and desire the benefits more than others who are eligible. This biases any

comparison of the response or performance of the two groups because their motivations and other characteristics were not the same.

Nonrandom comparison group methods--are commonly used when the requirements for strict randomized control cannot be satisfied. Attempts are made to make the comparison group as similar to the experimental group as possible by matching individuals with the same sex, age, racial, or socio-economic characteristics. The differences in results between the two groups (experimental and the matched comparison) are attributable, as in experimental designs, to the results of the program. However, without random assignment there is greater danger that the observed results are attributable to nonprogram influences. Other difficulties with the method include potential bias resulting from self-selection by participants.

Comparison of similar programs--attempts to establish measures and data with which the outcomes of two or more ongoing programs or components can be compared. Program comparisons are attractive for several reasons. They (1) provide information on effectiveness of alternatives in comparable terms and for the same time period, (2) reduce the need to rely completely on the elusive "control" of experimental methods applied to one project, (3) help generalize the results if widely distributed "representative" projects can be included, and (4) offer an opportunity to identify exceptional performance and to study what is operationally different about those projects.

Program comparisons comprehensive enough to yield the above advantages are costly and difficult to manage. For example, although "planned variations" must be carefully documented at the outset, once in operation they will seldom be free of further changes, which also must be documented. It should be noted whether such changes are "positive" (efforts to apply even better methods) or "negative" (resistance to adopting the prescribed methods).

Time series--involves a series of measurements at periodic intervals before the program begins and during the program. For example, in evaluating the safety results of Connecticut's crackdown on speeding, it was possible to use time series data collected for several years before and after this new policy change. An abrupt change in such trend data is strong evidence that the action taken caused the observed change in the trend. If measurements can also be obtained in another setting treated as a comparison group, additional insights are possible.

Careful interpretation is needed when using time series data. There may be a time lag between receipt of services and the impact of the services. The analyst should also be alert for cyclical phenomena, such as unemployment levels, which might cause part of the trend.

The methods discussed above are not exhaustive and there are other ways of making useful comparisons.

Developing needed information

Many information systems are not structured to capture data necessary for making comprehensive and valid appraisals. Consequently, a certain amount of ad hoc data collection will be necessary. Repeated appraisals of the same programs will be aided by incorporating procedures to capture the desired data on a continuing basis. For new programs, a special effort should be made at the beginning to incorporate specific provisions for gathering the necessary information. In any case, decisions are required on

--precisely what questions are to be answered and

--specific items of data required for analytical methods to be employed.

Selection, design, and implementation of data collection instruments may be the least attractive aspects of any appraisal, but they are among the most important. Major sources of data include

--interviews,

--mailed questionnaires,

--onsite observations,

--peer group ratings,

--standardized written tests,

--project and other program records,

--Federal and State government statistics, such as those from the Census Bureau and Bureau of Labor Statistics,

--performance tests or other physical evidence,

--clinical examination,

--financial, cost accounting, and operational management information, and

--documents such as minutes, progress reports, public releases, etc.

Usually, it is helpful to use several sources, and there are opportunities for creativity in design of collection instruments and in analytical designs which merge data from several sources. It may be very helpful to merge data obtained from personal interviews (condition of home, etc.) with data in program files (achievement scores, etc.). Careful design, in a technical sense, must be coupled with careful consideration of preserving the confidentiality of data about individuals. (See ch. 5.)

Interpreting program results

The key point in interpreting the data is to ascertain the degree to which results, consequences, or effects are attributable to the program(s) or to other external influences. Frequently the data will reveal only small effects. Even small effects are important, however, because they may be the only clue available to the potential for larger effects which were either obscured in the data or are achievable only through greater change in the program. Because of the potential for large effects to be obscured by the data, it is important to examine small effects very carefully.

Even if valid, dependable results are not obtained, the data should provide insight into the structure needed in further research and evaluation. These insights may also suggest new and different alternatives which should be assessed.

Those making appraisals have a responsibility to provide systematic information about the results of policies and programs and about the degree of confidence attached to these results. Where a high degree of uncertainty exists, it may preclude firm recommendations concerning policy and program actions. When recommendations are made in these circumstances, the uncertainty must be clearly communicated. Further appraisals can frequently reduce the uncertainties and provide a basis for firm recommendations.

Checking completeness of the appraisal

It is helpful in preparing an interpretive summary of a policy or program appraisal to view the interdependent concepts which have been discussed as a checklist.

Some of the questions which should be contained in such a checklist are:

--Were the reasons for the study found to be valid? Were the cause, scope, and intensity of the original problem or issue redefined as part of the study or as a result of the study? Why did it need attention at this time? Was full consideration given to the expressed needs of all potential users of the study?

--Were the objectives of the program or policy clearly identified? Did they shift over time? Were there implicit objectives?

--Were any special problems, either conceptual or practical, encountered in using input, process, output, efficiency, or effectiveness measures? Were valid standards for comparisons used? Was it necessary to employ surrogate measures and what was the rationale for their choice? What other quantifiable or intangible consequences were measured and how?

--Were data collection instruments sufficient under the circumstances?

--Are findings statistically significant and practically important? Do they answer questions posed at the beginning of the study?

--Were the hypotheses accepted? Were uncertainties resulting from problems with data identified and properly considered? Compared to other studies or evidence, do data and conclusions agree? If not, why not?

--Were the lessons learned identified? Can suggestions be made for immediate improvements?

--To what extent can the performance of this program be generalized to apply to other settings within which the program takes place or may take place? What should and should not be done in the future in other locations or in similar programs? Are these conclusions based

on demonstrated causal relationships? Are reasons for program weaknesses indicated?

--Have recommendations been developed for alternatives to be analyzed and compared?

--What is still left to be studied? What new questions were raised that require further research? Which areas of research still need further exploration? What research methods need to be developed or improved in order to make future appraisals more authoritative?

ASSESSING POLICY AND PROGRAM ALTERNATIVES

As in the case of appraising policy and program results, the methods used in assessing policy and program alternatives build on the fundamentals discussed at the beginning of this chapter. In this case also, there are additional concepts and methods which are needed, such as

--developing a range of alternatives,

--screening the preliminary alternatives,

--estimating the measurable consequences,

--assessing provisional orderings,

--determining the impact of constraints,

--reassessing the ordering of alternatives, and

--checking the completeness of the assessment.

Developing a range of alternatives

It is essential to search out a wide range of alternatives. The initial search for alternatives should not be constrained. Continuing, modifying, expanding, reducing, or abandoning an existing program should be included, as well as completely new alternatives. With regard to the existing program, consideration should be given to reexamining the validity of the existing objectives. The process of developing alternatives should include a thorough questioning of the need for any governmental intervention, which may have been justified on any of the following grounds:

--Absence of suitable private alternatives or absence

of a private marketplace in which the needed service can be distributed.

--The benefits to society resulting from universal use of services or facilities, such as sewage disposal.

--Equal availability of a service, such as public education.

--Distribution of benefits to disadvantaged people, such as health benefits through medicare and medicaid.

--Regulation of private activities, such as the certification of effectiveness and purity of drugs.

--Provision of incentives for desired private activities, such as development of energy resources.

Broad classes of approaches which show potential for solving the problem being analyzed should be initially identified. One or more promising alternative approaches from each of the broad classes should be developed. If broad classes are not examined, alternative approaches are usually unnecessarily limited to relatively small incremental changes from existing programs. For example, analysis of an incremental change in eligibility standards for the food stamp program is more narrowly defined than an analysis of overall income security or nutrition policy.

Reasonable alternatives from all sources, including those suggested by governmental agencies, legislative committees, advocacy or interest groups should be considered. Issue papers, such as described in chapter 5, can be useful at this stage of an assessment.

Screening the preliminary alternatives

A preliminary analysis of the likely consequences associated with the range of alternatives, including the status quo, should now be undertaken. This initial screening is intended to eliminate obviously inferior approaches and to reduce the original list of alternatives to a manageable size. It is helpful to make approximate calculations of cost and consequences, of break-even points, and of technical feasibility, etc. Alternatives should not initially be ruled out based on implementation difficulties, including organizational or procedural changes.

Modifications and combinations of alternatives usually become apparent and frequently provide the basis for new and superior alternatives. The search for alternatives is a continuing activity and the analytical effort provides opportunities to invent or discover other alternatives that will arise.

Estimating measurable consequences

Estimates must be made of anticipated measurable consequences as well as of all costs and resource inputs under various conditions and levels of available resources. Measurable consequences include effectiveness, side effects, and distribution considerations. In making such estimates, the data on actual costs and effectiveness found in prior appraisals of similar programs should be used together with actual operating data. It may also be necessary to use well developed causal models to make such effectiveness estimates. Although these models must adequately simulate the real situation, an existing model may serve. Experience has shown that it is costly and time consuming to develop a completely new model.

Some effort should be made to estimate side effects and their influence on resources. An estimate is needed, to the extent possible, of the differences of impact on the beneficiaries and the cost bearers (distribution considerations). Approximations may have to be used for side effects and distribution estimates, and various value judgments are involved in weighing both.

When analyzing costs which should be associated with effectiveness, various cost concepts are needed, and information on these costs is usually available. When analyzing costs which should be associated with side-effect and distribution considerations, total as well as incremental costs should be developed. Frequently such costs are incomplete. They should be checked for reasonableness and consistency with the alternatives of interest.

Information at the margin, as contrasted with information on total quantities, is very important in resource allocation decisions. Approximations of incremental costs, however, are more easily obtained than are approximations of the marginal aspects of other program consequences. A reasonable effort should be made to estimate the direction and magnitude of the variations of program consequences over relevant ranges.

Information on measurable consequences obtained from

audits, evaluations, or other studies should be used. Historical and trend data may provide information concerning how the various consequences are affected by the scale of activity.

Assessing provisional orderings

Once the total and incremental consequences of the alternatives have been estimated, the alternatives should be arrayed in some order. This ordering may be based on one of several available approaches.

One approach is "cost effectiveness." This approach focuses on resources expected to be consumed and how well the objectives are achieved. Using this framework, a preferred alternative is identified as one which produces the largest achievement for a given level of costs or which minimizes resources expended for attaining a given level of effectiveness.

While the cost-effectiveness approach provides a basis for ordering competing alternatives, it does not clearly allow for comparisons of alternatives associated with multiple, possibly conflicting, objectives. Other consequences of alternatives--side-effects and distribution considerations--are not an integral part of the analysis and may require separate examination.

A second approach to ordering alternatives is "cost-benefit" analysis. Side-effects and distribution considerations are incorporated in this approach. Major consequences, or benefits, are measured in dollars, and differences between monetary benefits and costs provide the basis for choice among alternatives. Theoretically, cost-benefit analysis is more useful than cost-effectiveness analysis in treating differing as well as conflicting objectives. The streams of benefits and costs can be discounted to their equivalent present values, thus accounting for the effects of time. Conceptually, decisionmakers could select programs based on rankings of net present value benefits (or derivatives of this data) until the total available resources were exhausted. However, this approach requires that all measures can be converted to dollars (a difficult task at best) in a way which the decisionmaker understands. Because of the difficulty of quantifying side-effects and distributional effects, there is often no clear distinction between cost-effectiveness and cost-benefit analysis.

Another approach is "cost-value" analysis. This is a technique for obtaining generally acceptable quantitative

weights for use in comparing the value of the alternatives. In this approach, the weights assigned to various outcomes are based on decisionmakers' judgments.

The cost-value method combines elements of cost-effectiveness and cost-benefit analysis. Side-effect and distribution considerations can be incorporated with effectiveness. Because the judgments of decisionmakers differ, various sets of judgments should be used and the ordering(s) of alternatives should be tested for its/their sensitivity to these differences. In such analysis both the array of consequences associated with each alternative and the ordering based on the various value systems should be presented to decisionmakers.

Each approach has both strengths and limitations, but all share certain limitations. One such limitation is uncertainty caused by such things as variations in assumptions and the quality of information on the alternatives. Because uncertainty is always present in anticipating future outcomes, undue reliance should not be placed on small differences in ordering(s) of alternatives. The quantitative analysis which has been discussed should be supplemented with an analysis of nonmeasurable consequences. A serious attempt should be made to indicate the significance of nonmeasurable consequences.

Determining the impact of constraints

Special efforts should be made to assess the impact of actual and potential legal, financial, and political constraints. Programs and policies must operate within the framework of law. Alternatives which may appear theoretically desirable must also operate within the law. Consequently, the alternatives considered for adoption must conform to this framework.

In addition to these sorts of constraints, there are constraints resulting from conflict with other objectives. An example of such constraints is the conflict between environmental, transportation, and energy objectives.

However, constraints are not inflexible. If decisionmakers were clearly aware of the potential opportunities foregone resulting from existing constraints, those constraints might change.

Decisionmakers must consider possible public reaction to alternative policy and program options, strategies that

might increase their acceptability, and what administrative or other operational barriers to implementation exist. The problems of implementation and of acceptability may, to some degree, be dealt with in analysis. Usually, assistance can be provided to decisionmakers in identifying the "second or third best" alternatives which may have higher prospects for being accepted or implemented.

It has been argued that if acceptability considerations are avoided, the assessment of alternatives becomes more objective, less parochial, and less tailored to fit preconceived positions. On the other hand, it may be argued that if acceptability considerations are not included, the analysis may prove to be irrelevant.

Reassessing the orderings of the alternatives

Orderings of alternatives are always provisional. They are determined within the context of the factors and values considered to be important during the course of the analysis. The assumptions and values underlying the various orderings of the alternatives must be clearly presented to decisionmakers. Furthermore, even when the analyst thinks the study is completed, decisionmakers may raise new issues, ask new questions, request further study, and ask for additional comparisons. As these requests are answered, the orderings of alternatives may shift.

Although attempts should be made to include as many factors as possible, other considerations properly affect the final policy and program choices. Some of these considerations may be completely beyond the analyst's knowledge or ability to estimate, even qualitatively, and belong in the province of the decisionmakers' judgments. However, the analyst should attempt to understand these considerations and to devise sensitivity analyses which may help to sharpen the decisionmakers' judgments. Analysts may even suggest new alternatives which balance achievement of conflicting objectives in ways not perceived when the initial set of alternatives was developed with the decisionmakers, or they may identify alternatives which keep options open or avoid irreversible damage or risk. Experienced analysts will seldom attempt to use an "optimizing" technique for this sort of communication with decisionmakers because many of the important considerations are neither specific enough nor quantifiable. Both analysts and decisionmakers must be satisfied with what, in their judgment, is a "good" but not necessarily the "theoretically best" alternative.

Checking completeness of the assessment

Some of the questions which should be considered in preparing an interpretive summary of a policy or program assessment are contained in the following checklist:

--Were the reasons for the study found to be valid? Were the cause, scope, and intensity of the original problem or issue redefined as part of the study or as a result of the study? Why did it need attention at this time? Was full consideration given to the expressed needs of all potential users of the study?

--Were the objectives of the program or policy explicitly stated and validated? Did they change during the course of the assessment? If so, why?

--Were there any potentially interesting alternatives eliminated early in the analysis? If so, why? Under what circumstances might they become attractive?

--Were any special problems, either conceptual or practical, encountered in specifying an adequate set of quantifiable measures? How reasonable were the dollar values attributed to physical measures, if that was done? Were qualitative indicators properly identified and used?

--Do the effectiveness measures accurately reflect the degree of attainment of the objectives? Were they consistently used among all of the alternatives? Is the effectiveness data reliable? Has uncertainty in the data been properly considered?

--Were side-effects and distribution considerations adequately considered? Are there significant differences among the alternatives?

--Were all of the cost implications captured? How reliable are they? What is the range of uncertainty?

--To what assumptions or data is the ranking of the alternatives sensitive? Are there any actions which can make the leading alternatives significantly less affected by the uncertainties?

--Are there any special problems connected with gaining

general acceptance of the apparently preferred alternatives? Will implementation of any of these pose particular difficulties?

--Is it likely that additional information about the leading alternatives would change the ranking? How, when, and at what cost could this information be obtained? Can the policy or program decisions be held open while new studies, evaluations, or research efforts are completed? What long-term evaluation or research efforts need to be initiated to meet similar or related problems in the future?

CHAPTER 5

PRACTICAL ASPECTS OF MANAGING AND PERFORMING STUDIES

Previous chapters have discussed the conceptual aspects of appraising the results of policies and programs and of assessing alternative solutions to the issues related to them. In deciding which problems to study and in carrying out actual evaluations or analyses, however, certain practical questions arise which need to be addressed. This chapter discusses some of these practical aspects and offers suggestions for coping with them. The list is not extensive but is indicative of the very real problems faced in this type of work.

FORMULATING AN AGENDA OF STUDIES

One of the most important responsibilities facing any manager of an evaluation, analysis, audit, or other program review staff is developing an overall work plan for the organization. Planning a program of studies which will be of maximum benefit to decisionmakers should involve two principal tasks:

--Identifying problems or issues which are evolving as major areas of concern.

--Deciding which of the many candidate problems the organization should commit itself to studying.

Identifying emerging problems

A contribution can be made to resource allocation decisions by raising problems and exploring their ramifications in "issue papers." These focus on problems which, there is reason to believe, will become the subject of a full-scale evaluation or analysis.

The ability to recognize emerging problems for issue papers depends on experience and good judgment. Developing an issue paper also requires an understanding of the problem area and its environment.

An issue paper may follow the format and style appropriate to a full-scale evaluation or analysis but is limited to an assessment of what is known about the problem. An issue paper could be as short as a few paragraphs or long enough to cover all or almost all of the points required in a full evaluation or analysis but without the scope or definitiveness of a

finished study. An issue paper should emphasize recommendations on the nature of further study efforts. For example, whether the problem should receive high priority attention (and why), whether it should be pursued but on a long-term basis (and why), or whether it should be abandoned (and why).

Deciding which problems to study

Many problems, programs, and policy issues are in need of systematic study. Yet, scarce staff resources need to be allocated to the most productive projects.

Issue papers can identify policy and program problems worth evaluating or analyzing. However, a complex series of judgments is still needed to select that particular group of problems which, if solved, would maximize the anticipated payoff. While it is relatively easy to list the factors influencing these choices, it is seldom feasible to appraise all of them in a formal quantitative fashion. In some cases there can be little more than an informed guess about the potential utility of a study.

A systematic weighing of the following factors will be helpful.

1. The anticipated payoff of successful evaluation or analysis.

This payoff can take several forms: an ineffective program can be canceled and costs saved; a mismanaged program can be reshaped with consequent improvements in effectiveness, reductions in costs, or both; or better alternatives can be substituted for current programs and policies with gains in effectiveness, reductions in cost, or both.

2. The chance of the successful performance of an evaluation or analysis.

This judgment depends on a basic understanding of the fundamental causal relationships; the requirements for additional information; the adequacy of current analytical methods; the quality of staff, consultants, or contractors; and the time and money available.

3. The chance that a preferred course of action can actually be implemented.

This judgment depends upon such things as newness, simplicity, visibility, coverage, and timeliness of the preferred course of action.

4. The need for resolving the problem or issue.

This need depends on the nature and relative importance of the problem and the time remaining before a meaningful decision has to be made.

5. The estimated cost of the evaluation or analysis.

BEGINNING A STUDY

Certain tasks need to precede major commitments of staff and other resources. These tasks include preparing a study plan, obtaining necessary agreements, selecting the study team, establishing lines of communication, and selecting appropriate methods.

Preparing a detailed study plan

Substantial effort should be devoted to drawing up a comprehensive and thorough study plan which will serve as a guide for all subsequent work. A study plan that is too broad in scope or loosely stated is almost certain to create false expectations for some interested groups. Clearly, trade-offs have to be made between the time devoted to planning versus doing a study and, within the planning period, between a detailed and a general study plan.

As the study progresses, it is likely to deviate from original expectations. Perhaps, the issue turns out to be different from that originally postulated; the objectives may not have been stated precisely enough; a working assumption may not prove viable; other alternatives to the program emerge; new facts come to light; hoped for data cannot be obtained; and so on. All of these developments call for some modification of the study plan. Changes should be made, as appropriate, to the study plan.

Essential elements of the study plan would appear to be:

--A clear statement of the problem to be studied, questions to be answered, and decisions to be affected.

--A careful listing of constraints and assumptions.

--A statement of methods to be used.

--A specification of the resources to be committed (including identification of the key staff members and any contracted tasks required).

--The frequency, format, and recipients of reports.

--Procedures for amending the study plan.

--The timeframe for the major components of the study and the final deadline.

When a study or a major part of it is to be performed by contract, there should be discussion and understanding by the parties concerning the essential elements of the study plan. This is likely to require lengthy dialogue with the decision-makers. Persons with official responsibility for the policy or program and for the study should assess feasibility and validity of the study plan. Any differences should be resolved before the study begins. Substantial time and effort may be necessary to arrive at a workable understanding. In the case of contract studies, the agency staff must be technically competent to oversee the study and must also be familiar with the various rules on contract management.

Selecting the study team

Most analyses or evaluations require contributions from several key persons. For large studies, subteams for particular aspects may be required. As in any group effort, someone must be in charge to (1) provide guidance, (2) manage the work on a day-to-day basis, (3) report to higher authority, and (4) generally be responsible for meeting the terms of the study plan.

The coordinator or director should be experienced, with a technically sound but broad background, an instinct for the principal issues, and the leadership abilities that elicit from the team members their best efforts. It usually turns out that the team coordinator or director will have to be principal editor of the final report--so writing skill is necessary.

A team studying any complex policy or program should be composed of experienced persons from various disciplines, with the stature required to obtain needed information and assure credibility of the study. Regardless of their origin, however, all should be made to feel as coequal members of an exciting intellectual experience and useful endeavor.

One way to create such an environment--at the same time avoiding duplication of effort--is to have an initial briefing on the terms of the study plan with all team members. Important aspects such as concepts, assignments, schedules, basic assumptions, need for personal and agency coordination, and reporting requirements should be fully understood and agreed upon in advance. Provision should be made for periodic briefings by each specialist to the team as a whole so that everyone has both a grasp of overall progress and a chance to offer facts or insights on any aspect of the study.

It is often helpful to obtain reviews by competent and widely recognized independent professional analysts and evaluators and experienced program administrators. This advice adds a seasoned viewpoint which may improve the technical aspects and may assist the supervisor in assessing the technical adequacy of the work of staff members trained in different disciplines.

Establishing lines of communication

If the study effort is sufficiently large, official points of contact among various interested groups and users of the study should be designated. This should insure that communications of all kinds flow quickly and clearly among the groups having a major interest in the progress of the study. Open communications provide the basis for a more complete assessment or appraisal and a climate in which recommended changes are more likely to be accepted and implemented.

Selecting appropriate methods

Analytical methods which yield valid and (hopefully) unequivocal results should be used. However, the method must also satisfy the constraints of time, money, and data peculiar to the study. If the constraints imposed are so rigid that the study would be compelled to use methods judged to be analytically inappropriate, the study should be undertaken only after fully informing responsible authorities of the risk that reliable conclusions and recommendations are not likely.

No particular approach or technique is inherently the appropriate one. In practice, there are too many attempts to mold the policy or program issue to fit a specific technique. This should be avoided. For a specific study, various approaches, each having its own particular logic should be considered. Usually, a blend of methods and techniques will be required to provide insights into the full consequenses

of the various alternatives. Reasons for selecting a particular approach or blend of approaches should be clearly stated so others can understand the rationale for the particular choice.

Whatever approaches and methods are selected, they should satisfy the following criteria:

1. Validity--how much confidence is there that the results can be actually used?

2. Relevance--are the results useful to decisionmakers? Will the method be used (such as a model) to answer decisionmakers' questions?

3. Significance--will the results go beyond what is apparent from direct observation? Will the results tell the decisionmaker something new and important?

4. Efficiency--does the value of the insights exceed the cost of using the approach?

5. Timeliness--will the analytical information be available in time to meet a management or legislative decision point such as renewal of expiring legislation?

Modeling and statistical inference are two related methods which are particularly useful and are frequently used by evaluators and analysts.

A model is an abstraction from or a representation of the key elements in some real system. If the key elements and their relationships are adequately specified, relevant, and valid, a model can predict the consequences of untried alternatives and variations in data and assumptions.

Statistical inference techniques are widely used to analyze data obtained from the various collection instruments and analytical models. However, conditions and assumptions underlying these methods must be satisfied if the method is to be used. Mistakes can occur, for example, if prepackaged computer programs are used without understanding the assumptions and conditions.

CONDUCTING A STUDY

In the performance of any evaluation or analysis, practical decisions of many types must be made, and practical problems are frequently encountered. Some of the most common ones are discussed here.

Collecting relevant data

In performing studies, there is often a temptation to collect all of the information which might be of use. While every piece of information may have some value in the right place, is it relevant and worth what it costs to acquire it? Questions which should be continually applied to any data collection effort are:

--Exactly what question is this piece of data intended to answer?

--What analytic model demands it?

--What calculation cannot be done without it?

Testing the reliability of data

An attempt should be made to estimate whether data is reasonable at the time it is first generated; i.e., how does this new piece of data square with everything else that is known or can be deduced relating to it? This is especially important when complex calculations are involved. How does the answer compare with rough calculations? The exercise of making rough calculations frequently gives the staff member new insights into the data.

There are numbers of one kind or another which are widely published. Everyone seems to use them unquestioningly. However, a careful analysis has often demonstrated that some data has a different interpretation than what is commonly supposed.

Occasionally, an attempt may be made to withhold information. It is not uncommon to hear that data

--is too hard to assemble,

--does not exist in the form wanted,

--is only a working paper, or

--is privileged.

When faced with this type of situation, the analyst should (1) consider the value of the information to the study, (2) attempt to obtain a release of the appropriate information if needed, and (3) propose to the study coordinator that a formal request be sent for the needed information. In some cases, essential data will have to be "constructed" or "extracted" from secondary sources.

Frequently, data collected from different sources about the same subject matter will be in apparent conflict. The first practical step in getting the right data is to reconcile the apparent conflicting interpretations of the data. An appropriate question may be: Are they truly two different sets of values describing exactly the same event or situation? A second step would be to examine how the data was derived. The apparent conflict may be a simple function of the data collection methodology. After these procedures have been employed, it may be appropriate to use an analytical technique to determine the significance of the differences. Additional assurances may be obtained by having data reviewed by experts in the field.

Protecting the confidentiality of information about individuals

It is often necessary in evaluation and analysis to collect data about individuals. In most research involving human subjects there has been a firm commitment to protect the confidentiality of personal data. It is important to make certain that data on individuals is not personally identifiable in the study or in unsecure files. If it is necessary to obtain information from the same individuals in subsequent time periods, special controls and procedures should be required to assure that systems of records do not disclose individually identifiable data.

Federal agencies and some Federal contractors are required to comply, where applicable, with all provisions of the Privacy Act of 1974 to protect the confidentiality of individually identifiable data. These provisions include

--public disclosure of the fact that an agency maintains a system of records about individuals,

--strictly enforceable procedures for assuring that individuals have access to their records and the opportunity to correct them,

--controls on disclosure of individuals' identifiable data, and

--administrative, technical, and physical safeguards to prevent unauthorized access to such data.

In planning a study, care should be taken to require individual identifiable information to be collected only when no other approach can enable the issue to be validly studied. When such data is collected, it must be properly protected.

Documenting and referencing

Documenting appraisals of results and assessments of alternatives is important. The documentation should be sufficient so that another individual or team involved in reviewing the policy or program, by reviewing the documentation, could follow the analysis and, as needed, reconstruct parts of it or use it in another study. Basic assumptions should be clearly identified and recorded. The rationale for using direct or surrogate measures should be stated explicitly. Oral interviews should be summarized in writing, dated, and filed. Original documents should be retained. Complete files of relevant raw data and work papers should be kept and filed so that they can be retrieved easily for review. Information which cannot be readily filed should be adequately described and referenced in the files.

The study team should design, use, and save work papers. Well designed, clearly labeled, and fully legible work papers offer an important insurance policy to the study team. The work papers constitute the evidence gathered. A review of the work papers will show whether the study team has been thorough or whether they may have overlooked an important fact or element of a problem and that all similar elements of the analysis or evaluation have been treated consistently. The work papers should be checked against the study plan to assure that the plan was carried out or that changes are fully explained. Developing the total costs of each of a series of alternatives is an outstanding example of the need for, and usefulness of, a carefully designed and clearly labeled set of worksheets. Without them, the chances of missing an important cost element, incorrectly calculating an intermediate result, or costing the competing alternatives inconsistently are substantial.

Work papers should be dated and signed so that a clear trail is established as to who did what and when. The best

way to tie it all together is to file, with workpapers, one copy of the final report which is cross referenced to significant sections of the workpapers.

Adhering to time schedules

Effort should be made to anticipate some of the possible delays, and the time schedule should allow for unforeseen delays. Most complex tasks are harder than originally anticipated and, therefore, take longer than estimated. In complex studies, detailed schedules for component parts may be necessary. A proposal to expand the scope of the study or to do more work in order to sharpen the results should be carefully justified, particularly if it involves risk of delay in the schedule.

Leading and coordinating the study team

It is essential to maximize the interaction among the study team members. Physical arrangements which inhibit this should be avoided or modified if at all possible. When gathering the first list of alternatives or hypotheses, brain storming is extremely useful.

The coordinator should take every practicable step to insure easy access to the decisionmakers who expect to use the analysis or evaluation. A continuing (but not necessarily continuous) dialogue should help to make the products useful and well accepted. The coordinator also needs to impress on the team the importance of maintaining an open, honest, and amicable relationship with the personnel of the program under analysis or evaluation. It is all too easy for program people to frustrate a study if they have been antagonized or hurt.

Using computer-based models

For most large-scale, but routine, quantitative manipulations (statistical analysis, linear programing, etc.) good "canned" programs are available and should be used. When a program or problem has many complex interrelationships, however, and the effects of altering the assumptions or data are not obvious, a specially designed, computer-based model may facilitate the study. In such cases, creative computer programers are extremely valuable.

The structure and operation of any model should be reasonably apparent to decisionmakers who want to use the study: its output and workings must be readily understandable to

them. Usually, this can be accomplished by carefully diagraming the components of the model and explaining how each component operates and interacts with the others. Users of the study will normally accept the computational competence of the model if the logic makes sense to them and they have confidence in the study team.

COMMUNICATING STUDY RESULTS

Many persons doing studies fail to understand that doing a good piece of work is necessary but hardly sufficient for bringing about a favorable change in the world. At least two major steps beyond successful completion of a study are required: the results must be clearly, concisely, and cogently communicated to all those affected and a policy or program decision must be made which results in some kind of action.

Specifying the nature of reports

There are three general classes of problems involved in reporting appraisals of results and assessments of alternatives: (1) to whom reports should be made, (2) when reports should be made, and (3) what style and content characterizes good reports. Each new study will suggest its own individual requirements and should be made a matter of record in the agreed work plan adopted before each study is begun. A few general guidelines can, however, be set down.

Obviously, the final report should be addressed to those who are in a position to take appropriate action--or to assure that it is taken by others. Unless special considerations dictate otherwise (e.g., security problems), reports should routinely go first to the team supervisor and others as needed to insure that they meet the organization's professional standards. Even professionally sound studies, however, may result in disagreements with the managers of the programs being studied. In these cases, the study team should reduce the number of areas of disagreement; and, where these continue to exist, the issues should be substantial and clearly defined. Although decisionmakers waiting to use the report should be kept informed of key findings, it will in the end serve them best if the review process is complete before the final report goes to them.

No report, other than the final version, should be distributed beyond those mentioned above without their concurrence. Unauthorized release of preliminary, draft, interim, or partial

reports can be harmful because, frequently, erroneous information, even though corrected later, becomes widely diffused and becomes a source of further error and confusion. Publicly available reports should be free of such errors.

In planning the study, sufficient time should be allowed for writing the final draft report, gathering comments, editing, and securing the necessary approvals. The report writers, in turn, have an obligation to complete the report within the scheduled time. Report outlines should be prepared early. They can provide indications of the most critical data gathering and interpretation tasks yet to be completed in order to have a useful and timely report. Decision points come and go relentlessly, and a potentially good, decision-affecting report may lose much of its value because it was not available when needed.

Communicating with clarity and conciseness

Writing a good report is an art, and the required skills are probably as scarce as those necessary for evaluation and analysis. The solution is to insist that staff members work at learning to write well. One helpful step is to provide staff members with specific guidance, such as a good style manual, and insist that they study and use it as part of their regular duties. In addition, someone on the staff can serve as resident editor. It is frequently helpful to have a skilled technical editor or writer join the team when the report is being written. All significant alterations should be discussed with the author: not only to insure accuracy, but also to assist the author in learning to write shorter and more trenchant reports.

Study reports are typically directed at a reader lacking relevant technical training. Therefore, the main body of the report should be written so that it is readily comprehensible to the nonprofessional reader. However, material included in the report should be sufficient so that a reader can understand the arguments in support of the conclusion. Jargon should be kept to a minimum, and where it is used, define it carefully. Supporting technical material should be presented in appendixes. Graphs and tables included in the main body of the text must be clearly labeled and fully discussed in the text. Short reports are typically self-contained, while long ones ought to be accompanied by an executive summary of the study's general conclusions and recommendations.

There will, of course, be differences between the format and content of a report on appraising program or policy results

and a report on assessing alternatives, or a report containing both. Within each of these, some variation in format and content is inevitable, depending on the nature of the policy or program issue being studied and the methods used. In general, the format and content of reports should cover what was found through each of the concepts and methods discussed in chapter 5, as appropriate.

Following up

Writing a clear, concise, and informative "final" report is not the end of the "communicating" responsibilities. Usually, some decisionmakers will need assistance in (1) interpreting the report, (2) clarifying aspects of it, (3) getting answers to questions raised by it but not answered, and (4) in general, developing a reasoned reaction to it. Briefings, informal question and answer sessions, and various kinds of supplementary written materials may be needed. In some cases, the communicating responsibility may even extend to preparing the supporting technical parts of whatever document emerges from the decisionmaking process.

It is the responsibility of the staff which performed the work to be available to the decisionmaker to help in understanding and using the study. The staff should also make a diligent effort to find out whether or not the study was useful. Lessons learned in this way can lead to better studies the next time.

AN ANNOTATED BIBLIOGRAPHY

The purpose of this appendix is to list some references that should be useful to those persons having limited experience in conducting evaluations and analyses. The references are listed under several categories: basic disciplines, quantitative methods, evaluation, and analysis.

BASIC DISCIPLINES

Baumol, William J., Economic Tneory and Operations Analysis. 3d ed. Englewood Cliffs, N.J., Prentice-Hall, Inc., 1972.

An excellent treatment of basic economic concepts and quantitative methods as they would apply to issues of resource allocation.

Downs, Anthony, Inside Bureaucracy. Boston, Little, Brown, 1967.

Very useful insights concerning bureaucratic behavior and setting.

Haveman, Robert H. and Julius Margolis, eds., Public Expenditures and Policy Analysis. Chicago, Markham Publishing Co., 1970.

Various aspects of public expenditure economics are discussed. The economic bases of public expenditures are developed in part I. Part II, with its emphasis on institutional considerations, is of special interest, as is part III which is concerned with analytic problems in policy analysis. The remaining parts survey the planning, programing, budgeting, and system experience and offer suggestions.

Rivlin, Alice M., Systematic Thinking for Social Action. Washington, D.C., The Brookings Institution, 1971.

A provocative series of essays on the issues involved in attempting to solve the problems of society.

Strunk, William and E. B. White, The Elements of Style. New York, Macmillan, 1959.

A very good writing style manual stressing clear and concise writing.

QUANTITATIVE METHODS

Hillier, Frederick S. and Gerald J. Lieberman, Introduction to Operations Research. San Francisco, Holden-Day, Inc., 1967.

Although a knowledge of mathematics is required, the text presents a comprehensive survey of the methods, models, and techniques that are used in analyses.

Levin, Richard I. and Charles A. Kirkpatrick, Quantitative Approaches to Management. 2d ed. New York, McGraw-Hill, 1971.

An introduction to quantitative methods and techniques.

Moroney, M. J., Facts From Figures. 3rd ed. Baltimore, Penguin Books, Inc., 1956.

A very readable treatment of the use and misuse of statistical techniques.

Raiffa, Howard, Decision Analysis: Introductory Lectures on Choices Under Uncertainty. Reading, Mass., Addison-Wesley Publishing Co., 1968.

A clear exposition of the process of determining best choices under uncertainty and of considerations affecting group decisions.

Tanur, Judith M. and others, eds., Statistics: A Guide to the Unknown. San Francisco, Holden-Day, Inc., 1972.

Applications of statistics and probability are developed in a case method setting.

Wonnacott, Thomas H. and Ronald J. Wonnacott, Introductory Statistics. 2d ed. New York, John Wiley & Sons, Inc., 1972.

A fairly rigorous, but understandable text of statistical inference, including Bayesian methods and nonparametric statistics.

EVALUATION

Hatry, Harry P., Richard E. Winnie, and Donald M. Fisk, Practical Program Evaluation for State and Local Government Officials. Washington, D.C., Urban Institute, 1973.

A very useful primer on evaluation with good examples, including an excellent discussion of comparison methods.

Isaac, Stephen and William Michael, Handbook in Research and Evaluation. San Diego, Knapp, 1971.

A compendium of useful checklists, do's and don'ts, and summaries of important concepts and techniques for evaluation.

Riecken, Henry W. and Robert F. Boruch, eds., Social Experimentation: A Method for Planning and Social Intervention. New York, Academic Press, 1974.

An invaluable guide--technically, ethically, and administratively--in using experimental designs for evaluations. Excellent annotated bibliography on experiments.

Suchman, Edward A., Evaluative Research: Principles and Practices in Public Service and Social Action Programs. New York, Russell Sage, 1967.

Considered to be a classic text in evaluation. Chapter IV, Categories of Evaluation, gives a framework useful in developing evaluative questions for a proposed study.

Weiss, Carol, Evaluation Research: Methods of Assessing Program Effectiveness. Englewood Cliffs, N.J., Prentice Hall, 1972.

An excellent introductory text for evaluation.

Weiss, Carol, ed., Evaluating Action Programs: Readings in Social Action and Education. Boston, Allyn and Bacon, 1972.

Collection of articles dealing with basic concepts and issues in evaluation, especially for social programs.

ANALYSIS

Amacher, Ryan C., Robert D. Tollison, and Thomas D. Willett, eds., The Economic Approach to Public Policy. Ithaca, N.Y., Cornel University Press, 1976.

A collection of articles discussing the role of economists in public decisionmaking and applications in a range of programs, problems, and issues.

Dolbeare, Kenneth M., ed., Public Policy Evaluation. Beverly Hills, Sage Publications, 1975.

A collection of recent thoughts on distinctions between social science research and policy analysis related to political feasibility and action, with extensive discussion of applications to crime control.

Dorfman, Robert, ed., Measuring Benefits of Governmental Investment. Washington, D.C., The Brookings Institution, 1965.

A series of contributed papers concerned with the application of cost-benefit analysis. Wide ranging applications are discussed.

English, J. Morley, ed., Cost-Effectiveness. New York, John Wiley & Sons, Inc., 1968.

A series of papers covering the various aspects of cost-effectiveness analysis.

Fisher, Gene H., Cost Considerations in Systems Analysis. New York, American Elsevier Publishing Co., Inc., 1971.

A comprehensive treatment of all relevant cost considerations that should be treated in analysis. Chapters 1, 2, 3, and 6 are especially useful.

Gregg, Phillip M., ed., Problems of Theory in Policy Analysis. Lexington, Massachusetts, Lexington Books, 1976.

An excellent source of recent thought on theoretical problems and premises of policy analysis.

Hatry, Harry, Louis Blair, Donald Fisk, and Wayne Kimmel, Program Analysis for State and Local Governments. Washington, D.C., Urban Institute, 1976.

A practical, readable primer on analysis, including the basic elements of program analysis.

Hitch, Charles J. and Roland N. McKean, The Economics of Defense in the Nuclear Age. Cambridge, Harvard University Press, 1960.

Although the application is military, a classic discussion of cost-effectiveness is found in chapters 7, 9, 10, 11, and 12.

Nagel, Stuart S., ed., Policy Studies and the Social Sciences. Lexington, Massachusetts, Lexington Books, 1975.

A collection of articles emphasizing the contribution to policy studies which can be made by particular disciplines, including sociology, psychology, anthropology, mathematics, law, and political science.

Quade, E. S., Analysis for Public Decisions. New York, American Elsevier Publishing Co., Inc., 1975.

A highly important and readable book on all analytic aspects involved in formulating and implementing policy decisions.